

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 08-287188

(43)Date of publication of application : 01.11.1996

(51)Int.Cl. G06K 9/62
B07C 3/14
G06K 9/20
G06K 9/34
G06K 9/42

(21)Application number : 07-090584

(71)Applicant : NEC CORP

(22)Date of filing : 17.04.1995

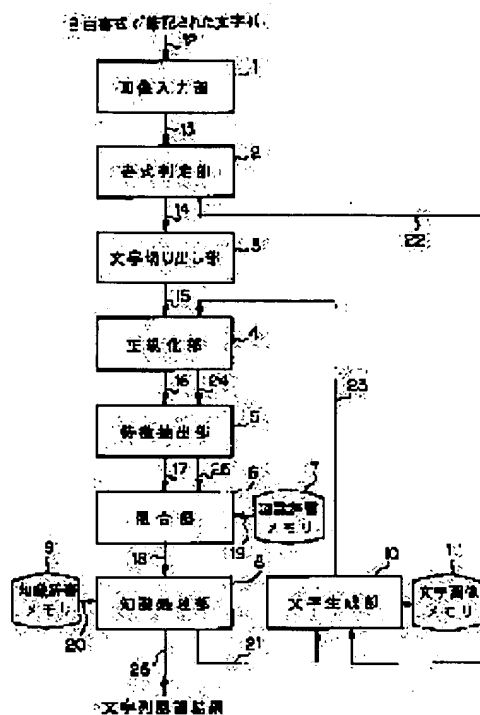
(72)Inventor : YAMAUCHI TOSHIFUMI

(54) CHARACTER STRING RECOGNITION DEVICE

(57)Abstract:

PURPOSE: To provide a character string recognition device which can automatically recognize a character string, entered in free format with a character touch and entanglement, fast with high precision.

CONSTITUTION: This device has an image input part 1, a format decision part 2, a character cutting part 3, a normalization part 4, a feature extraction part 5, a collation part 6, a knowledge processing part 8, a character generation part 10, a 1st means, and a 2nd means. At this time, the 1st means can be equipped with the normalization part 4, feature extraction part 6, and collation part 6. Further, the 2nd means can be equipped with the knowledge processing part 8. Recognition candidate characters 18 obtained by collating feature patterns 17 of an input character with reference patterns 19 by the collation part 6 are matched by the knowledge processing part 8 against knowledge dictionary data 20, and a character category 21 which is estimated to be desirable for an inconsistent part is outputted. A character generation part 10 generates a character image 23 corresponding to the character category 21 and the input character image 15 is matched again against the generated character image 23.



LEGAL STATUS

[Date of request for examination] 17.04.1995

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 2751865

[Date of registration] 27.02.1998

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

JAPANESE

[JP,08-287188,A]

CLAIMS DETAILED DESCRIPTION TECHNICAL FIELD PRIOR ART EFFECT OF THE INVENTION TECHNICAL
PROBLEM MEANS OPERATION EXAMPLE DESCRIPTION OF DRAWINGS DRAWINGS

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Industrial Application] It is related with the approach of carrying out automatic recognition of the character string notes of was taken with free formats (indeterminate pitch), such as the address, a name, etc. written to the character string written to the document which does not have the limit on entry of a letter face etc. especially, or mail, about the approach of carrying out automatic recognition of the character string written to a document or mail.

[0002]

[Description of the Prior Art] When the alphabetic character which read the document with the limit on entry of a letter face etc., and adjoined in target character string recognition equipment contacts, it is few, and logging of an alphabetic character is comparatively easy. on the other hand, contact of the alphabetic character which adjoined in the character string recognition equipment of the free format which recognizes the address, the name, etc. written to a document without the limit on entry of a letter face etc., or mail and alphabetic characters -- becoming intricate -- etc. -- it is difficult to presume the field of one character correctly, and many cases where alphabetic character logging is difficult occur. Therefore, it is necessary to make it correspond to contact of these alphabetic characters, or the difficulty of alphabetic character logging depended for becoming intricate in the character string recognition equipment for a free format.

[0003] There are the following two approaches as the character string recognition approach of the conventional free format.

[0004] The 1st approach is the approach of searching for two or more kinds of combination of logging of a character string, and choosing the combination of suitable logging by adjustment with knowledge information. This approach is "alphabetic character logging and recognition" (Murase others) from the handwriting character string using language information. The Institute of Electronics and Communication Engineers paper magazine, 86/9, Vol.J69-D, No.9, pp1292-September, 1986 [1301 or], And it is indicated by "handwriting character string recognition using alphabetic character connection information" (an Institute of Electronics, Information and Communication Engineers technical report besides Kobayashi, PRU-91-67, pp39-October, 1991 46 or]).

[0005] It specifically asks for the basic block which started the character string to hyperfractionation, and the combination of two or more kinds of alphabetic character logging is acquired as a combination of a basic block. Next, individual character recognition is performed about the integrated block (a basic block should put together) corresponding to the combination of each alphabetic character logging. In the knowledge based system section, the knowledge information about the recognition candidate alphabetic character which is the output of the individual character recognition corresponding to each integrated block, and the alphabetic character for reading is adjusted at the last, a candidate with the highest adjustment is read, and it considers as a result.

[0006] Although processing whose 2nd approach starts an alphabetic character from a character string is performed only about a general candidate, it is the approach of registering the description data of the alphabetic character pattern which contacted or became intricate beforehand as a recognition dictionary about the character string with possibility that contact of an alphabetic character or alphabetic characters become intricate, collecting per word of two or more characters, and performing reading.

[0007]

[Problem(s) to be Solved by the Invention] In the 1st approach, a right logging result may not necessarily be included in the phase of asking for the basic block which started the character string to hyperfractionation.

[0008] For example, when "size" and a "mound" of a character string "Otsuka" touch, an alphabetic character can be recognized, it can judge that the character string consists of two characters for the first time, and it becomes possible to start as two characters. However, since it is asking for the basic block from the character string before recognizing an alphabetic character in the 1st approach, when the character string touches, it is difficult [it] to search for the point starting [compulsive]. Therefore, the combination which started "size" and a "mound" cannot necessarily be extracted as a basic block. In this case, a right logging result will be included.

[0009] Thus, although the probability which includes a right logging result by making the number of basic blocks increase in the 1st approach can be raised, it is difficult to set a probability including a right logging result to 1 completely. Moreover, when the number of basic blocks is made to increase, the number of the integrated blocks which are the combination of a basic block increases exponentially. For this reason, performing individual character recognition about all integrated blocks, and performing a knowledge based system about the combination of each integrated block further has the trouble that throughput will become very large.

[0010] Moreover, in the 2nd approach, it is not realistic to register the combination of all the alphabetic characters that contact. When three or more characters contact especially, there is a trouble that the number of combination of an alphabetic character becomes huge, and a number of registration becomes huge.

[0011] This invention aims at offering a high speed and the character string recognition equipment which can carry out automatic recognition with high precision for the character string contact of an alphabetic character and alphabetic characters become intricate, and notes of is taken with the ***** free format in view of such a point.

[0012]

[Means for Solving the Problem] The image input section into which the character string recognition equipment of this invention inputs the image of the character string for [notes of is taken with the free format] reading, The format judging section which judges the form information which extracts a string area out of said inputted image, and shows the direction of whether a character string is columnar writing or it is lateral writing, The alphabetic character logging section which cuts down an alphabetic character pattern in the predetermined number of alphabetic characters from the image of said string area extracted by said format judging section, The normalization section which performs normalization processing according to a character size to each of said alphabetic character pattern cut down by said alphabetic character logging section, The feature-extraction section which extracts the alphabetic character description of the normalization pattern of the input-statement character obtained by said normalization section, and generates the description pattern of an input-statement character, The collating section which performs matching between the description pattern of said input-statement character, and the reference pattern stored in recognition dictionary memory, and outputs a recognition candidate alphabetic character based on the result of matching, The knowledge based system section which outputs the alphabetic character category presumed to be desirable about the part which takes adjustment with the knowledge dictionary data of said recognition candidate alphabetic character and the alphabetic character for [which is stored in knowledge dictionary memory] reading, and cannot take adjustment, The alphabetic character generation section which generates the alphabetic character image corresponding to said alphabetic character category outputted to one alphabetic character category from said knowledge based system section using the alphabetic character image data stored in the alphabetic character image memory which has memorized one alphabetic character image, Perform normalization processing in said alphabetic character image generated by said alphabetic character generation section, and the normalization pattern of a generation alphabetic character is obtained. 1st means to extract the alphabetic character description of the normalization pattern of this generation alphabetic character, to obtain the description pattern of a generation alphabetic character, to collate the description pattern of this generation alphabetic character, and the description pattern of said input-statement character, and to obtain a recognition candidate character string, Adjustment with said recognition candidate character string obtained as a collated result and said knowledge dictionary data is taken, and it has 2nd means to output a character string with the highest adjustment as a character string recognition result.

[0013] As for the character string recognition equipment of above-mentioned this invention, said 1st means can be equipped with said normalization section, said feature-extraction section, and said collating section.

[0014] Moreover, as for the character string recognition equipment of above-mentioned this invention, said 2nd means can be equipped with said knowledge based system section.

[0015]

[Function] thus, after constituted this invention performs recognition processing of an input-statement character, only about the alphabetic character which cannot take adjustment Since the alphabetic character image with which two or more characters touch, the alphabetic character image which divided one character, or the alphabetic character image which combined both is generated and it collates with an input-statement character image from the alphabetic character category presumed to be desirable It becomes possible a high speed and to carry out automatic recognition with high precision about the character string contact of an alphabetic character and alphabetic characters become intricate, and notes of is taken with the ***** free format, without performing useless processing.

[0016] Moreover, when three or more characters touch, or even when the alphabetic character is located in a line with the longitudinal direction, it becomes possible to generate an alphabetic character image and to collate with an input-statement character image, and it becomes possible a high speed and to carry out automatic recognition with high precision about the character string contact of an alphabetic character and alphabetic characters become intricate, and notes of is taken with the ***** free format.

[0017]

[Example] Below, the example of this invention is explained with reference to a drawing.

[0018] Drawing 1 is the block diagram showing the configuration in one example of this invention. In this example, the case where recognition processing is performed per one character is explained.

[0019] The configuration of drawing 1 is explained. Drawing 1 has the composition of having the image input section 1, the format judging section 2, the alphabetic character logging section 3, the normalization section 4, the feature-extraction section 5, the collating section 6, the recognition dictionary memory 7, the knowledge based system section 8, the knowledge dictionary memory 9, the alphabetic character generation section 10, and the alphabetic character image memory 11.

[0020] The image input section 1 inputs the image of the character string 12 notes of was taken with the free format of the document for reading, or mail, and performs processing which generates image data. Image formation of the reflected light of the light irradiated from the light source is specifically carried out with image sensors, such as CCD (Charge Coupled Device), this is scanned, and it changes into an electrical signal. Next, the whole surface image data 13 of multiple-value level is generated by quantizing the changed electrical signal.

[0021] The format judging section 2 searches for the rectangle description, the histogram description of the axis of ordinate and axis of abscissa of a black pixel, etc. to circumscribe about the black pixel lump in the whole surface image data 13 generated in the image input section 1. And while judging the form information 22 which shows the direction of columnar writing and lateral writing of a character string according to distribution of these descriptions, the string area image 14 is generated.

[0022] The alphabetic character logging section 3 presumes an one-character field in quest of the rectangle description, the histogram description of the axis of ordinate and axis of abscissa of a black pixel, etc. to circumscribe about the string area image 14 generated in the format judging section 2, and performs processing started as an input-statement character image 15. However, in alphabetic character logging processing of the character string 12 notes of was taken with a free format like the address and the name on a document without a letter face, or mail, an one-character field cannot necessarily be presumed correctly. The case where the field of two or more characters is accidentally presumed as an one-character field as a general example, and a field smaller than one character may be accidentally presumed as an one-character field. The input-statement character image 15 in this example is defined as an image of the field which presumed as an one-character field and was started. Therefore, if the one-character field may be started correctly, the logging error may have generated the input-statement character image 15.

[0023] The normalization section 4 performs normalization processing which arranges the alphabetic character image size of the input-statement character image 15 cut down in the alphabetic character logging section 3, and the generation alphabetic character image 23 mentioned later. As the approach of normalization processing of arranging image size, there are a linearity normalization approach expanded or reduced by ratios [longitudinal direction / a lengthwise direction or], the nonlinear normalization approach of changing the ratio expanded or reduced by the field, etc. In this example, it is possible to apply any normalization approach.

[0024] The feature-extraction section 5 extracts the alphabetic character description of the input-statement character image 16 which was generated in the normalization section 4 and which normalized, and the generation alphabetic character image 24 which normalized. As an alphabetic character description to extract, there are the direction description of the edge of an alphabetic character, an alphabetic character stroke consistency description, the structural-analysis-description, etc. In this example, it is possible to apply the feature-extraction approach of arbitration.

[0025] The collating section 6 performs matching between the description pattern 17 of an input-statement character image, and the reference pattern 19 registered into the recognition dictionary memory 7. The check of whether the description component belonging to a distance value, similarity, or the reference pattern 19 exists in the description pattern 17 of an input-statement character image, By the check of whether the description component belonging to the description pattern 17 of an input-statement character image exists in the reference pattern 19 etc., the degree of collating is judged and the recognition candidate alphabetic character 18 is outputted. This processing is defined as the 1st collating processing.

[0026] By considering study alphabetic character image data as an input, the recognition dictionary memory 7 performed normalization processing by the normalization approach used in the normalization section 4, and has registered the description pattern extracted by the feature-extraction approach used in the feature-extraction section 5, and the category code pair to which study alphabetic character image data belongs as a reference pattern 19.

[0027] The knowledge based system section 8 puts the recognition candidate alphabetic character 18 in order based on the location of the input-statement character image 15 in the inside of the string area image 14 by using as input data the recognition candidate alphabetic character 18 called for in the collating section 6, and generates a recognition candidate character string. The knowledge database for reading is beforehand registered into the knowledge dictionary memory 9. Although there is various knowledge, such as the address, a name, a firm name, a trade name, and a components name, as a knowledge database, the case where an address knowledge database is used is explained in this example. In the knowledge based system section 8, a recognition candidate character string is searched based on the knowledge data 20 registered into the knowledge database for reading.

[0028] As a result of performing alphabetic character logging in the alphabetic character logging section 3, the normalization in the normalization section 4, the feature extraction in the feature-extraction section 5, collating in the collating section 6, and the knowledge based system in the knowledge based system section 8, the alphabetic character category 21 considered to be desirable from the knowledge based system section 8 about the part which cannot take the knowledge data 20 for reading and adjustment is outputted.

[0029] The alphabetic character generation section 10 generates the alphabetic character image corresponding to the alphabetic character category 21 outputted from the knowledge based system section 8 based on the alphabetic character image memorized by the alphabetic character image memory 11 and the form information 22 outputted from the format judging section 2, and outputs it as a generation alphabetic character image 23.

[0030] The generation alphabetic character image 23 outputted from the alphabetic character generation section 10 is inputted into the normalization section 4, and the generation alphabetic character image 24 which normalized is obtained. Next, the feature extraction of the generation alphabetic character image 24 which normalized in the feature-extraction section 5 is performed, and the description pattern 25 of a generation alphabetic character image is obtained.

[0031] Next, in the collating section 6, collating processing is performed again. Matching between the description pattern 17 of an input-statement character image and the description pattern 25 of a generation alphabetic character image is performed. The check of whether the description component belonging to the description pattern 25 of a distance value, similarity, or a generation alphabetic character image exists in the description pattern 17 of an input-statement character image, By the check of whether the description component belonging to the description pattern 17 of an input-statement character image exists in the

description pattern 25 of a generation alphabetic character image etc., the degree of collating is judged and the recognition candidate alphabetic character 18 is outputted. This processing is defined as the 2nd collating processing.

[0032] Moreover, the collating section 6 has a means to compare the 1st collating processing result with the 2nd collating processing result. This is a function of the degree of adjustment and the description pattern 17 of an input-statement character image of the description pattern 17 of an input-statement character image, and the good reference pattern 19 of most the adjustment in the recognition dictionary memory 7, and the description pattern 25 of a generation alphabetic character image which compares the degree of adjustment and outputs the good recognition candidate alphabetic character 18 of the degree of adjustment in the 1st collating processing. As the comparison approach of a concrete collating processing result, the with a distance value [1st] average is compared with the with a distance value [2nd] average, and there are an approach of outputting the recognition candidate alphabetic character 18 with the smaller average of a distance value, the approach of comparing the average of the 1st similarity with the average of the 2nd similarity, and outputting the recognition candidate alphabetic character 18 with the larger average of similarity, etc.

[0033] The knowledge based system section 8 uses as input data the good recognition candidate alphabetic character 18 of the degree of adjustment which compared the 1st collating processing result with the 2nd collating processing result, and was obtained in the collating section 6. It searches based on the knowledge data 20 which read a recognition candidate character string and are again registered into the target knowledge database, and the result which was able to take adjustment is most outputted as a character string recognition result 26 with the knowledge data 20.

[0034] Drawing 2 is drawing explaining the procedure in one example of this invention, and shows the case where the character string "Otsuka Shinden, Kawagoe-shi" is inputted. Moreover, the address database is used as knowledge for reading. In an address database, it is assumed as address data that there are "Suna-Shinden, Kawagoe-shi", "Amanuma Shinden, Kawagoe-shi", and "Otsuka Shinden, Kawagoe-shi." Moreover, in the recognition dictionary memory 7, it is assumed that there are 13 kinds of reference patterns 19 of a "river", "***", a "city", "sand", new ["new"], a "rice field", "heavens", a "swamp", "size", a "mound", "***", the "right", and "smallness."

[0035] The actuation in the configuration of drawing 1 is explained using drawing 2. In drawing 2, since the "mound" touches two characters in an input string "size", it is difficult to dissociate as two characters, "size" and a "mound", in the alphabetic character logging section 3, and the input-statement character image 15 is cut down as one character of "Otsuka." Therefore, the input-statement character image 15 generated in the alphabetic character logging section 3 becomes six alphabetic blocks 31-36.

[0036] In the character recognition section which consists of the normalization section 4, the feature-extraction section 5, and the collating section 6, it recognizes about six alphabetic blocks 31-36. In the collating section 6, collating processing is performed about the description pattern 17 of an input-statement character image, and 13 kinds of reference patterns 19 of the recognition dictionary memory 7, and the recognition candidate alphabetic character 18 the "Kawagoe **** recently-developed paddy field" is obtained. However, although there are the name of a places, such as "Suna-Shinden", "Amanuma Shinden", and "Otsuka Shinden", in the address of Kawagoe-shi in an address database, there is no name of a place of "*****." Therefore, it is presumed with the knowledge data 20 of the knowledge based system section 8 that the alphabetic character category 21 considered to be desirable as an alphabetic character of an alphabetic block 34 is "sand", "Amanuma" or, and "Otsuka."

[0037] The alphabetic character generation section 10 generates the alphabetic character image corresponding to a character code. The generation alphabetic character image 37 with which "heavens" and a "swamp" are perpendicularly located in a line, and the generation alphabetic character image 38 with which "size" and a "mound" are perpendicularly located in a line are obtained from the information that it is columnar writing of the format judging section 2. Since it is collating ending in the 1st collating phase, it is not necessary to generate again and to collate about the one-character image of "sand."

[0038] Next, in the normalization section 4, the generation alphabetic character images 37 and 38 are normalized, and the generation alphabetic character images 39 and 40 which normalized are obtained. In the feature-extraction section 5, the feature extraction of the generation alphabetic character images 39 and 40 which normalized is carried out, and the description pattern 25 of a generation alphabetic character image is obtained.

[0039] Then, in the collating section 6, 2nd collating processing between the description pattern which normalized and carried out the feature extraction of the alphabetic block 34, and the description pattern of the generation alphabetic character images 39 and 40 which normalized is performed. Moreover, in the collating section 6, the 1st collating processing result is compared with the 2nd collating processing result, and a result with the sufficient degree of adjustment is obtained as a recognition candidate alphabetic character 18.

[0040] In drawing 2, rather than adjustment with the reference pattern "***" whose adjustment between the generation alphabetic character images 40 and alphabetic blocks 34 which normalized "Otsuka" which he is 2nd as a result of collating processing is 1st as a result of collating processing, it judges with adjustment being good and "Otsuka Shinden, Kawagoe-shi" is obtained as a recognition candidate alphabetic character 18.

[0041] In the knowledge based system section 8, since the address data "Otsuka Shinden" exist in an address database, "Otsuka Shinden, Kawagoe-shi" is obtained as a character string recognition result 26.

[0042] Drawing 3 is drawing explaining the procedure in one example of this invention, and shows the case where the character string "Suna-Shinden, Kawagoe-shi" is inputted. Moreover, as knowledge for reading, the address database is used like drawing 2. In an address database, it is assumed as address data that there are "Suna-Shinden, Kawagoe-shi", "Amanuma Shinden, Kawagoe-shi", and "Otsuka Shinden, Kawagoe-shi." Moreover, in the recognition dictionary memory 7, it is assumed that there are 13 kinds of reference patterns 19 of a "river", "***", a "city", "sand", new ["new"], a "rice field", "heavens", a "swamp", "size", a "mound", "***", the "right", and "smallness."

[0043] Drawing 4 is drawing showing the data in the alphabetic character image memory in the configuration of drawing 1. As shown in drawing 4, the alphabetic character category 80 and the alphabetic character image data 83 are matched and stored in the alphabetic character image memory 11. Moreover, it is also possible to add the attribute information 81 and 82 to each alphabetic character image data 83-87. In drawing 4, the case where the attribute information 81 and 82 which shows whether the alphabetic character image data 83-87 has separated into the lengthwise direction or it has separated into the longitudinal direction is added is shown. For example, although it separates into two partial patterns of "MU" and "RO" in a lengthwise direction, since a "base" does not separate into a longitudinal direction, it becomes = (length, width) (1 0). However, although it separates into two partial patterns of "Si" and "***" in a longitudinal direction, since a "swamp" does not separate into a lengthwise direction, it becomes = (length, width) (0 1). Moreover, although it is also possible as these attribute information 81 and 82 to add a vertical separation location, a horizontal separation location, etc. further, as shown in drawing 4, the case where the partial pattern has been arranged so that the separation location 86 may be simply set as the location of the one half of one character is explained.

[0044] The actuation in the configuration of drawing 1 is explained using drawing 3. In drawing 3, since one character in an input string "sand" has separated into two alphabetic blocks and two alphabetic blocks separate and exist, it is difficult to unify as one character of "sand" in the alphabetic character logging section 3, and the input-statement character image 15 is cut down as two characters, the "right" and "smallness." Therefore, the input-statement character image 15 generated in the alphabetic character logging section 3 becomes seven alphabetic blocks 51-57.

[0045] In the character recognition section which consists of the normalization section 4, the feature-extraction section 5, and the collating section 6, it recognizes about seven alphabetic blocks 51-57. In the collating section 6, collating processing is performed about the description pattern of the input-statement character images 66 and 67 and 13 kinds of reference patterns 19 of the recognition dictionary memory 7 which normalized. Collating processing with the reference patterns 64 and 65 of "size" and a "mound" and the input-statement character images 66 and 67 which normalized is shown in drawing 3. As a result of collating processing, the reference pattern with the "right", the input-statement character image 67 which normalized, and most adjustment sufficient [the good reference pattern of the input-statement character image 66 which normalized, and most adjustment] serves as "smallness", and the recognition candidate alphabetic character 18 "Kawagoe right smallness Nitta" is obtained. However, although there are the name of a places, such as "Suna-Shinden", "Amanuma Shinden", and "Otsuka Shinden", in the address of Kawagoe-shi in an address database, there is no name of a place of "**** Nitta." Therefore, it is presumed with the knowledge data 20 of the knowledge based system section 8 that the alphabetic character category 21 considered to be desirable as an alphabetic character of alphabetic blocks 54 and 55 is "sand", "Amanuma" or, and "Otsuka."

[0046] The alphabetic character generation section 10 generates the alphabetic character image corresponding to a character code. About the two-character image of "size", a "mound" and "heavens", and a "swamp", since it is collating ending in the 1st collating phase, it is not necessary to collate again. Since it is not collating about the one-character image of "sand", two generation alphabetic character images "the stone" and "***" which divided generation alphabetic character image "sand" 59 into the longitudinal direction are obtained from the information that it is lateral writing of the format judging section 2.

[0047] Next, in the normalization section 4, two divided generation alphabetic character images are normalized, and the generation alphabetic character images 60 and 61 which normalized are obtained. In the feature-extraction section 5, the feature extraction of the generation alphabetic character images 60 and 61 which normalized is carried out, and the description pattern 25 of a generation alphabetic character image is obtained.

[0048] Then, in the collating section 6, 2nd collating processing between the description pattern which normalized and carried out the feature extraction of the alphabetic blocks 54 and 55, and the description pattern of the generation alphabetic character images 60 and 61 which normalized is performed. Moreover, in the collating section 6, the 1st collating processing result is compared with the 2nd collating processing result, and a result with the sufficient degree of adjustment is obtained as a recognition candidate alphabetic character 18.

[0049] In drawing 3, rather than adjustment with the reference pattern "the right" whose adjustment between the generation alphabetic character images 60 and 61 which normalized the "stone" which it is 2nd as a result of collating processing, and "***", and the input-statement character images 66 and 67 which normalized is 1st as a result of collating processing, and "smallness", it judges with adjustment being good and "Suna-Shinden, Kawagoe-shi" is obtained as a recognition candidate alphabetic character 18.

[0050] In the knowledge based system section 8, since the address data "Suna-Shinden" exist in an address database, "Suna-Shinden, Kawagoe-shi" is obtained as a character string recognition result 26.

[0051]

[Effect of the Invention] It has a high speed and the effectiveness that automatic recognition can be carried out with high precision for the character string contact of an alphabetic character and alphabetic characters become intricate, and notes of is taken with the ***** free format, without performing useless processing by generating an alphabetic character image from the alphabetic character category presumed that it is desirable only about the alphabetic character which cannot take adjustment after this invention performs recognition processing of an input-statement character, as explained above, and collating with an input-statement character image.

[0052] Moreover, when three or more characters touch, or even when the alphabetic character is located in a line with the longitudinal direction, an alphabetic character image can be generated, and it can collate with an input-statement character image, and has a high speed and the effectiveness that automatic recognition can be carried out with high precision for the character string contact of an alphabetic character and alphabetic characters become intricate, and notes of is taken with the

***** free format.

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] The block diagram showing the configuration in one example of this invention

[Drawing 2] Drawing explaining the procedure in one example of this invention

[Drawing 3] Drawing explaining the procedure in one example of this invention

[Drawing 4] Drawing showing the data in the alphabetic character image memory in the configuration of drawing 1

[Description of Notations]

- 1 Image Input Section
- 2 Format Judging Section
- 3 Alphabetic Character Logging Section
- 4 Normalization Section
- 5 Feature-Extraction Section
- 6 Collating Section
- 7 Recognition Dictionary Memory
- 8 Knowledge Based System Section
- 9 Knowledge Dictionary Memory
- 10 Alphabetic Character Generation Section
- 11 Alphabetic Character Image Memory
- 12 Character String Notes of was Taken with Free Format
- 13 Whole Surface Image Data
- 14 String Area Image
- 15 Input-Statement Character Image
- 16 Input-Statement Character Image Which Normalized
- 17 The Description Pattern of Input-Statement Character Image
- 18 Recognition Candidate Alphabetic Character
- 19 Reference Pattern
- 20 Knowledge Data
- 21 Alphabetic Character Category
- 22 Form Information
- 23 Generation Alphabetic Character Image
- 24 Generation Alphabetic Character Image Which Normalized
- 25 The Description Pattern of Generation Alphabetic Character Image
- 26 Character String Recognition Result
- 31, 32, 33, 34, 35, 36 Alphabetic block
- 37 38 Generation alphabetic character image
- 39 40 Generation alphabetic character image which normalized
- 51, 52, 53, 54, 55, 56, 57 Alphabetic block
- 58, 62, 63 Character string of a knowledge based system result
- 59 Generation Alphabetic Character Image
- 60 61 Generation alphabetic character image which normalized
- 64 65 Reference pattern
- 66 67 Input-statement character image which normalized
- 80 Alphabetic Character Category
- 81 82 Attribute information
- 83, 84, 85, 87 Alphabetic character image data
- 86 Separation Location

[Translation done.]